32ND RESEARCH SYMPOSIUM ASNT | EVENTS.

Presented At: ASNT Research Symposium:
25 – 28 June 2024
Pittsburgh, Pennsylvania, USA

Transfer Learning Mask R-CNN for Radiograph Image Quality Indicator Localization

Merrill Dennison[1], Connor Seavers[1], Tsuchin Chu[1],

[1] Southern Illinois University
1230 Lincoln Drive
Carbondale, IL, 62901
(618) 453-7049; fax (618) 453-7658; email mead.dennison@siu.edu

ABSTRACT

Ensuring the quality of industrial radiography is integral to its use for the inspection of components. This research work focused on applying an existing object detection and instance segmentation framework called Mask R-CNN to the recognition of image quality indicators (IQIs) in industrial radiographs. For the purposes of this project, a Mask R-CNN has been trained in a supervised manner, beginning with pre-trained ImageNet weights, toward the identification and localization of IQIs within digital radiographs. The goal of training the Mask R-CNN is for it to learn a mapping from the input radiographs to the output predictions of the bounding boxes and masks for each IQI in the input radiographs. On a high level, the Mask R-CNN serves as a function that takes a digital radiograph as input and provides a Python dictionary object as output. The output dictionary contains each region of interest predicted by the model for the given input radiograph, as well as their class IDs, probability scores, and mask images. Mask R-CNN is shown to be capable of adequately segmenting IQIs from radiographs when the standard practices for IQI placement are followed. This study explored the difference in Mask R-CNN performance when the training datasets are both small and contain IQIs at various orientations. A comparison is made between models trained with only parallel IQI examples and models trained with parallel, transverse, and askew IQI examples. This publication is focused on providing readers with a general understanding of the concepts, and numerical results are omitted in favor of visually depicting the best Mask R-CNN predictions. The results of this study have important implications for the application of existing computer vision and narrow artificial intelligence systems toward the detection of quality assurance objects within industrial radiography.

Keywords: industrial radiography, digital radiography, image quality indicators, supervised deep learning, Mask R-CNN

INTRODUCTION

Radiography is a nondestructive evaluation (NDE) technique that is commonly used in the aerospace industry to locate internal flaws in both ferrous and nonferrous materials. Irrespective of the material caught between the detector and the X-ray source, the quality of each radiograph must be adequately assessed, and image quality indicators (IQIs) are used towards this end. In digital radiography, image quality is most often analyzed using quantitative parameters such as signal-to-noise ratio (SNR), contrast-to-noise ratio (CNR), and image unsharpness [1]. An IQI is defined as a device or combination of devices whose demonstrated image or images provide visual or quantitative data, or both, to determine radiographic quality and sensitivity [2]. This research work is primarily focused on the segmentation and localization of plaque hole-type IQIs in digital radiograph images. Plaque hole-type IQIs are most often used for the determination of the CNR for a given radiograph and for qualifying the sensitivity of the radiograph according to industrial standards and guidelines based on the CNR value. Obtaining metrics related to the quality and sensitivity of the radiographic technique naturally follows the recognition of the IQI within the radiograph. This research work focused on applying an existing deep learning model for object detection and instance segmentation, called Mask R-CNN, to the problem of finding the plaque hole type IQIs as an initial step towards measuring the CNR in the IQI and determining its acceptability. The progression of the paper begins with some background information and a discussion of the methodology employed. Then, the results will be presented and discussed, followed by a section concluding the paper.

BACKGROUND AND METHODOLOGY

Object detection and instance segmentation are closely related image processing tasks that entail the leveraging of features within the digital images to perform a narrow intelligent task. Within the context of IQI recognition, object detection involves drawing a bounding box around each IQI and labeling it as such. Instance segmentation entails the classification of every pixel in the image as belonging to one of the IQIs that are present. Mask R-CNN [3] is capable of performing both tasks at once. The differences in these tasks are illustrated in Figure 1 below. With object detection, the interest is in obtaining bounding box coordinates and their corresponding class labels, while instance segmentation involves attributing a class label to every pixel in the image.

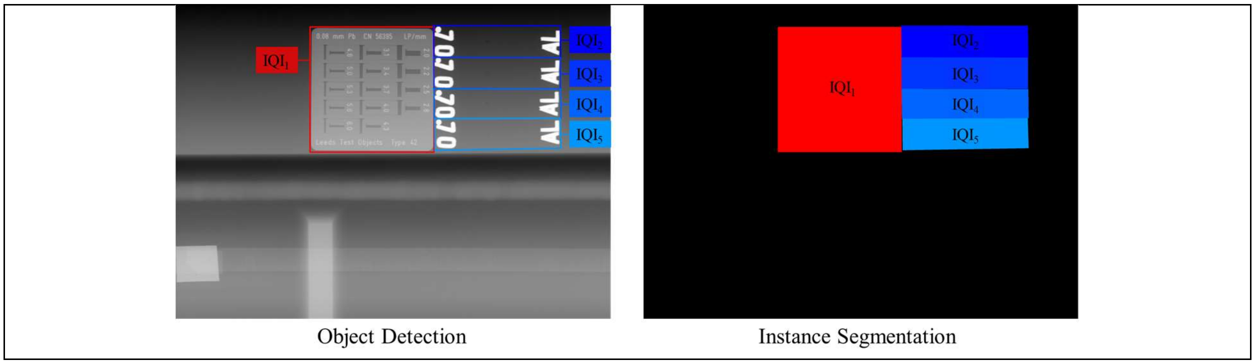

Figure 1: Example of object detection versus instance segmentation

Mask R-CNN is a state-of-the-art deep learning framework developed by the Facebook AI Research group in 2017 that has participated in deep learning and image recognition competitions that involved hundreds of distinct class categories. Mask R-CNN can detect objects within an image and generate a high-quality segmentation mask for each instance [3]. Figure 2 provides a diagrammatic description of Mask R-CNN where the input image is fed forward through the network, and a set of bounding boxes is produced as an output. Each output bounding box has a classification (confidence) score and a mask associated with it. The confidence score of a trained model is a measure of how confident it is that an IQI is present in the bounding box region. The utility of such a model lies in the potential for it to generalize well to images that were not in its training set, reducing the time required to generate accurate labels for each IQI in a very large set of radiograph images.

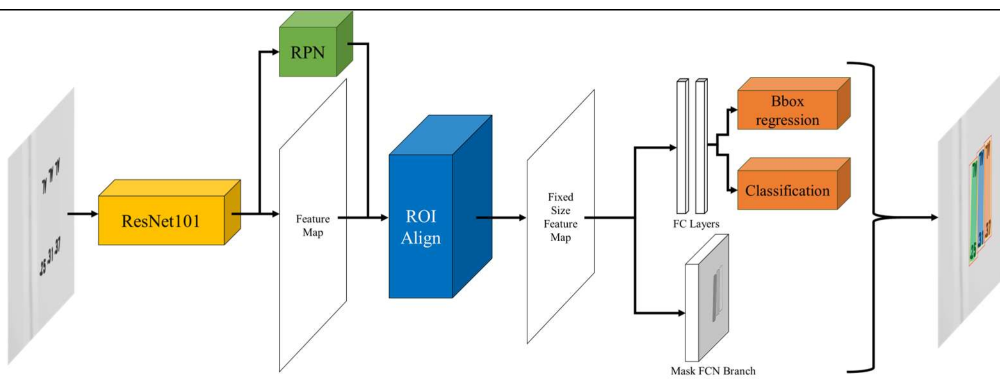

Figure 2: Mask R-CNN diagram

Mask R-CNN is trained via supervised learning and is comprised of several components whose exact mathematics could be the subject of an entire paper of its own. Supervised learning is a method of model training that involves the supplementing of input-output pairs to the model with iterative updates that bring the model's prediction closer to the output for every input. The model weights are initialized from a repository of pre-trained ImageNet weights, and thus, the supervised learning is technically called transfer learning, as the weight values obtained from pretraining on the ImageNet competition dataset are directly transferred to the model prior to the start of the supervised learning process. Supervised learning starts with a dataset of input-output pairs $\mathcal{D} = (\mathbf{x}_\mathrm{i}, \mathbf{y}_\mathrm{i})$ where each $\mathbf{x}_\mathrm{i}$ and $\mathbf{y}_\mathrm{i}$ are the i^{th} input and output, respectively. The dataset is then broken into training and testing portions making the dataset $\mathcal{D} = \{(\mathbf{x}_\mathrm{m}, \mathbf{y}_\mathrm{m}), (\mathbf{x}_\mathrm{n}, \mathbf{y}_\mathrm{n})\}$ where m is the dummy index for examples in the training set and n is the dummy index for the testing set examples. It is desired that the mask R-CNN serve as a function mapping all inputs to their corresponding output pairs. It is the mathematics of backpropagation that enables the model to be incrementally updated in the direction of the correct mapping. This also means that a model trained in a supervised manner on a finite dataset is optimized for making predictions on that finite dataset and may not necessarily generalize to input examples extraneous to that finite dataset used for training. As the ImageNet competition dataset does not contain examples of IQIs, the weights pre-trained on it cannot necessarily be expected to generalize well to the problem at hand. Every prediction provided by the model $(\hat{\mathbf{y}})$ is either correct or incorrect. The model is considered "trained" when the difference between $\mathbf{y}_\mathrm{m}$ and $\hat{\mathbf{y}}_\mathrm{m}$ is minimized for all m and the model is considered to be "generalized" when the difference between $\mathbf{y}_\mathrm{n}$ and $\hat{\mathbf{y}}_\mathrm{n}$ is minimized for all n.

A GitHub repository [4] of the Mask R-CNN model running on Tensorflow 2.0 was utilized within this research work. The goal of Mask R-CNN in this project is to output a bounding box and mask label for every IQI present in a radiograph image. To achieve this, radiographs with IQIs had to be procured and then labeled with mask, bounding box, and class information. The approach used in this work was to label the radiographs utilizing a custom Python program that started with the definition of four points within the radiograph for every IQI. These four points complete a four-sided polygon for each IQI and are used to generate mask images. Mask images start out as blank images of the same shape as the original radiograph. The mask generation process involves assigning a unique greyscale value to all pixels lying inside a given polygon region for all polygons. Once the mask has been generated, bounding box information is determined for each polygon region. So, a polygon is defined for each IQI, and for each polygon, a bounding box. An XML file containing the filename, bounding box, and class information is generated for every image. With the filename, bounding box information, and path to the mask image from the XML file, the radiograph images and their labels are ready to be opened and manipulated within the Python environment. The primary performance metric used to assess the model's ability to generalize to the images in the test set is the mean average precision (mAP). Readers interested in learning more about this metric and how it is used with Mask R-CNN are referred to an existing publication [4] on the subject. It suffices, for the presentation of results in this publication, for the reader to know that increasing mAP implies that the model made better predictions, while a mAP closer to zero means that the model made more inaccurate predictions with respect to the labels provided to it.

RESULTS AND DISCUSSION

A single training and inference loop involves (a) generating the training, validation, and testing subsets, (b) training the model, and (c) obtaining the mAP value for the trained model on the testing subset. IQIs are present within the data at different orientations. The majority of examples fall under the "Parallel" orientation. The datasets used for training are broken up according to the IQI orientation and the augmentation applied. For each dataset, several training and inference loops are run with randomization of the training, validation, and testing subsets between each loop. The mAP on the testing subset is averaged across all runs. All augmentations are applied "on the fly" during training, and test images are not augmented prior to inference. The following are observations of the study:

- mAP declined with the introduction of augmentations for datasets containing only parallel examples.
- mAP for datasets with only parallel examples was higher overall.
- mAP increased with the introduction of augmentations for datasets containing nonparallel examples.

A visual depiction of 'good' test results is provided in Figure 3 below. These results are ones where the predicted bounding box, mask, and label matched their ground truth counterparts. A well-trained model would ideally make predictions just like these for every input image. The original radiograph images are very faint to the human eye and thus contrast enhanced originals are provided above the predictions for human viewers to see the IQIs.

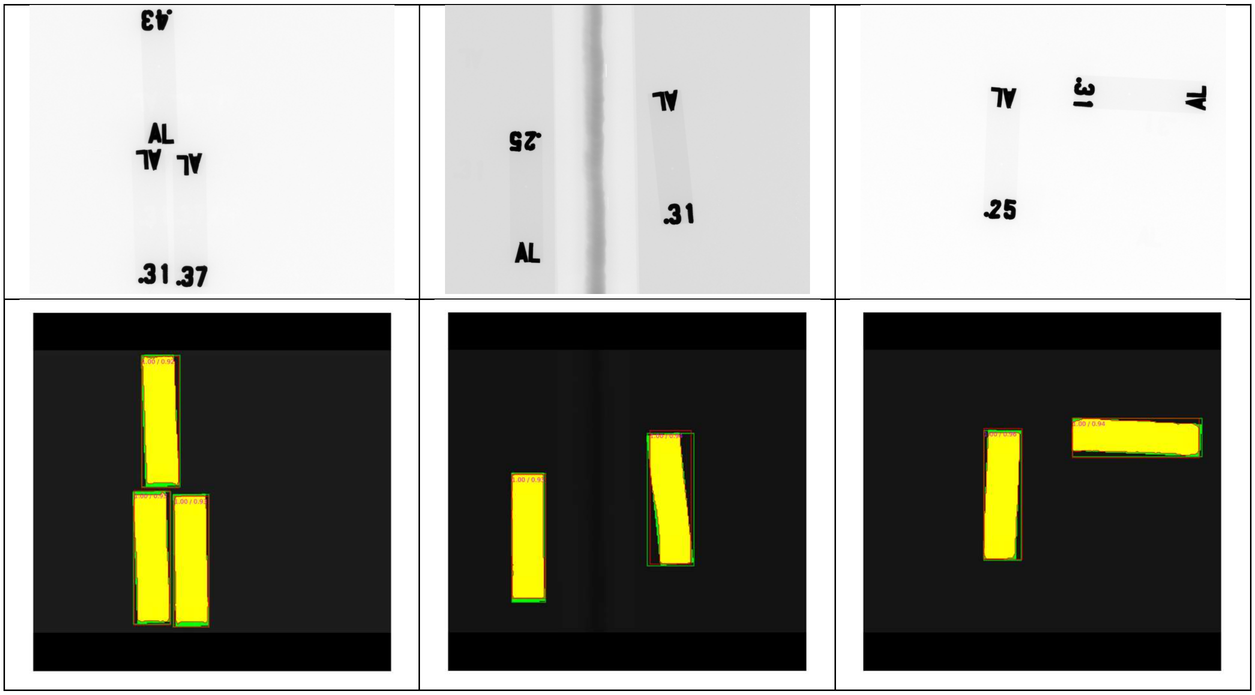

Figure 3: Example predictions made by a trained Mask R-CNN model

CONCLUSION

In this investigation, multiple instances of Mask R-CNN were trained, initialized from weights pretrained on the ImageNet competition dataset, to detect IQIs in industrial radiographs using a limited set of images. Training datasets were constructed based on the relative orientation of IQIs (parallel vs. non-parallel + parallel) and the number of augmentation operations applied during training. Results show that the Mask R-CNN generalizes better to the datasets containing only parallel instances of IQIs, while models trained on both parallel and non-parallel IQI examples saw a drop in performance. This is likely due to the relatively low number of non-parallel examples, making them outliers in each dataset. This scenario can be easily prevented in industrial radiographic applications, though, as ASTM E 2698-18e1 [5] calls for parallel orientation (± 5° relative to the weld) of IQIs in radiographs.

ACKNOWLEDGMENT

The authors would like to express their gratitude to Dr. Shant Kenderian and the Aerospace Corporation for their generous support and funding throughout the duration of this research project.

REFERENCES

(1) Khan, M., and Brothers, M., 2018, “Equivalent Penetrameter Sensivity for Performance Evaluation of Computed Radiography Systems,” *12th European Conference on Non-Destructive Testing (ECNDT 2018)*, Gothenburg, *e-Journal of Nondestructive Testing*, **23**(8). https://www.ndt.net/?id=22782
(2) 2024, ASTM E1316: Standard Terminology for Nondestructive Examinations.
(3) He, K., Gkioxari, G., Dollár, P. and Girshick, R.B., 2017, “Mask R-CNN,” *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2980-2988.
(4) Abdulla, W., 2017, “Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow,” https://github.com/matterport/Mask_RCNN
(5) 2018, ASTM E2698: Standard Practice for Radiographic Examination Using Digital Detector Arrays.